Towards Wearable Active Vision Platforms

W.W. Mayol, B. Tordoff and D.W. Murray
Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK
{wmayol,bjt,dwm}@robots.ox.ac.uk
http://www.robots.ox.ac.uk/ActiveVision

Abstract

This paper describes the design and construction of a wearable active vision platform which is able to achieve substantial decoupling of the camera motion from the wearer's motion. Design issues in sensor placement, robot kinematics and their relation to wearability are discussed and the prototype platform's performance is evaluated in a number of important visual tasks. The paper also discusses potential application scenarios for this kind of wearable visual robot.

1 Introduction

In the context of wearable computing, the active vision paradigm, pioneereed in work by Bajscy [1], Ballard and Brown [2], Aloimonos et al. [3] and others [4], has many of the advantages which are readily transferable to the wearable domain.

However, current work in wearable visual computing uses *passive* body-mounted cameras which make the imagery and image measurements dependent on the wearer's posture. Often it is assumed, or rather *hoped*, that the camera is pointing in the relevant direction by virtue of being mounted on the wearer's head. Even when pointing in the roughly the correct direction, any visual processing relying on feature correspondence from a passive camera is made more difficult by the large image displacements which arise when the wearer moves.

Rather than being fixed in the user's body frame, an autonomous wearable sensor may wish to make measurements in two further frames of reference: one centered on the stationary world, and the other centered on an independently moving object. A sensor that is fixed in one of these frames of reference will have difficulty making precise measurements in the other two.

We argue that optionally decoupling the visual sensor from the wearer is desirable. In section 2 we discuss in broad terms the design issues involved in wearable visual robotics before presenting in section 3 the design and implementation of the prototype wearable active vision camera shown in Figure 1. In section 4 we examine the performance of this prototype device in a number of visual tasks.

Figure 1: 3-axis Wearable Visual Robot. Closeup: 1) 2D Accelerometer. 2) CMOS Color Camera. 3) Elevation axis. 4) Pan axis. 5) Cyclotorsion axis.

2 Positioning the sensor

Positioning a camera on the human body is more problematic than that for mobile robots, as evidenced by the variety of solutions. Hat-mounted cameras have been used [5, 6] to look down at the user's hands and reaching space, whereas in [7] cameras are strapped to the wearer's hands themselves. In [8], the hat-mounted camera looks forward, an orientation also used when the camera is attached to a head mounted display [9]. In contrast, [10] uses a camera is worn on the chest. These placements are based on task feasibility and performance predicated on *passive* cameras.

If an active camera is used, we suggest that it is possible to consider more directly the issues of absolute field of view (FOV) and wearability.

A camera worn on the chest has the advantage of covering most of the user's working space, pointing to where the

user's handling/manipulative attention is, and is thus useful for user-centred applications. However our aim with wearable robots is to try to access both user and world-centered frames of reference, and this position is obviously limited by occlusion in the backwards direction and also laterally by movement of the arms. Another disadvantage of this position is when the user is seated in front of a desk the FOV is further restricted. Considering wearability, a chest camera might be too easily knock by the hands.

Placing the camera at the ear may be seen as a good alternative since it seems to have the largest range of FOV, and also has the benefit of being head-mounted and therefore looking where the user is looking. Although this position is useful for addressing the user-centered frame of reference, it actually complicates the decoupling of user movements. Wearability is also less than perfect: unrestricted and natural views of the face are important in social interaction.

In our work we have adopted a position on the shoulder, by attaching the camera to a collar loosely fitting around the neck. This position appears to be a good compromise between body-stable fixation, large virtual FOV and intrusion into the facial area of the wearer.

An alternative to an active camera might be the use of either a panoramic camera, or multiple cameras worn in different locations. Although panoramic cameras using mirrors and associated firmware to remap the image onto a plane [11] are now commercialized, the body of the wearer would occupy a unreasonable large fraction of the image. The use of a inherently passive panoramic devices does nothing to solve the large displacement correspondence problem mentioned earlier. This problem also remains unaddressed if a number of sensors were worn at different locations on the body. Indeed, using imagery from multiple cameras mounted on the flexible human form will increase the difficulty matching.

3 System Description

3.1 Kinematics

Figure 2 shows the configuration of axes used to provide elevation, panning and cyclotorsion (rotation about the camera's optical axis).

The kinematics follow a Helmholtz chain, 'elevate then pan', but unlike our other active heads, cyclotorsion is eliminated using a third motor. In *non-wearable* applications, stereo platforms usually follow the Helmholtz (or common-elevation) model so that the vergence geometry is simple. Monocular cameras however usually use the Fick chain of 'pan then elevate' to eliminate cyclotorsion about the optical axis. When the pan axis is kept vertical, the vertical is preserved in the image under these kinematics.

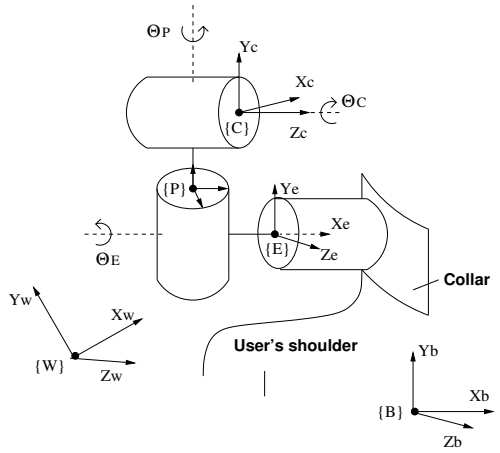

Figure 2: Coordinate frame of the three-axis Wearable Visual Robot. The camera's optic axis is aligned with Z_C.

In *wearable* applications, however, the pan axis will not remain vertical, and so an extra axis to eliminate cyclotorsion is essential, thus re-opening the choice of kinematic chain. Our choice of the Helmholtz chain is based on the likelihood of the body postures giving rise to kinematic singularities. Referring to Figure 2, if we perform a $\pm 90^\circ$ rotation in the Z_b axis, we obtain a singularity where elevation becomes pan and vice versa. This introduces tedious control-logic problems and the requirement of additional sensors to detect this state. If the platform is placed at the base of the neck, this singularity occurs for the Helmholtz configuration when the user lies on his side on a horizontal surface. If however the kinematics follow the 'pan then elevate' Fick chain, the singularity occurs when the user lies on his chest or back, a more stable and hence more likely configuration. Thus we adopt the Helmholtz configuration.

3.2 Actuators

The selection of actuation method for an active vision system do not depends purely on the physical properties of the actuator such as weight, volume or power consumption but also on the desired *behavior*. Currently, there are several actuation methods that may be used in the constrained-scale scenario imposed by wearable applications.

Some of the most promising methods include Shape Memory Alloy actuators (SMA), ultrasonic motors and the more traditional electromagnetic devices. The main advantage for the SMA approach is its high strength to weight ratio. This kind of actuators are basically suited for bang-bang control-like cases which do not match directly with the intermediate states and smooth movements required by an autonomous visual sensor. However, state of the art configurations shows that servo-control bandwidths of about 2 Hz are achievable with SMAs [12].

Another promising alternative uses ultrasonic signals to move a single rotor element at various degrees of freedom [13]. These kind of ultrasonic motors usually do not require mechanical amplificators nor reduction mechanisms and can keep a state without consuming power. However, by now, the devices remain bulky and the required high voltage and driving electronics complicate its use in our scenario of interest.

Properly geared miniature motors overcome the limitations of the above methods. They can produce smooth movements, require small voltage (3-6 V), moderate power consumption position (begin able to hold position without power), and have fast response, low weight and high torque; there is a wide range of controllers available to drive them.

The prototype's motors are lightweight servo motors combining motor, controller and gear head in some 6 cm^3 volume, producing axis torque of 0.06 Nm and axis speeds of 10 $rad.s^{-1}$. The embedded servo controller has the property of sending a command signal only if there is a position disturbance, and therefore effectively implement a minimum power management strategy of value in portable applications. The range of motion is of about $160°$ at each axis. The weight for each servo-device is about 6 gm.

3.3 Visual Sensor

A complete CMOS camera with sensors and driving circuitry can be embedded in a single chip reducing weight, size and power consumption, making CMOS cameras more attractive for wearable applications than traditional CCD-based sensors. The prototype's visual sensor is a CMOS colour camera with a field of view of some $50°$ and with a volume of about 7 cm^3. The camera is attached to the final axis of Figure 2 and weights 20 gm. Although using colour increases power consumption and volume in comparison with monochrome, colour clues are valuable in image segmentation.

The camera signal in the prototype is sent directly to the computer by the means of an umbilical connection. However, it is becoming straightforward to eliminate cables with the recent advances in the miniaturization of wireless video links operating in the GHz frequency range.

3.4 Inertial Sensors

Fixed to the camera is a two-axis accelerometer ADXL202 (Analog Devices) for gravity-vector tracking and/or sensing the user's motion.

As with conventional mobile robots, the inclusion of inertial sensing can simplify actuator control and save computational resources if properly fused with the other available sensors. In the case of an accelerometer able to measure dynamic and static acceleration, when no movement is present, the value acquired after digitization is related to the gravity vector by a linear relation of the form [17]:

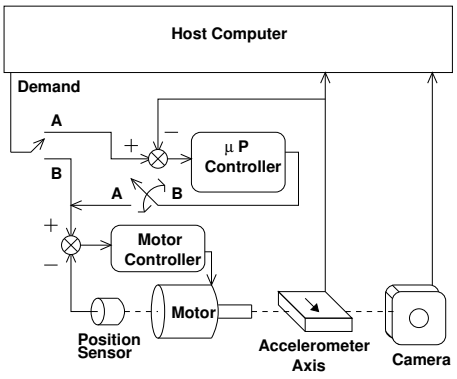

Figure 3: Control architecture.

$$A_i = O_i + (1 + J_i)\mathbf{g} \cdot \hat{\mathbf{n}}_i \quad (1)$$

where A_i is the signal from axis i, O_i the offset and $(1+J_i)$ the gain. The unit vector $\hat{\mathbf{n}}_i$ defines the direction of the accelerations along the given axis which is roughly known but has to be determined, and $\|\mathbf{g}\| = 9.81\ \mathrm{m.s}^{-2}$. As can be seen, A_i is maximum when $\mathbf{g}$ is aligned with $\hat{\mathbf{n}}_i$ and minimum when they are in the opposite directions.

Calibration methods are widely documented in the literature [17], however, since the sensor is used in a closed loop configuration (i.e. the sensor moves with the camera), accurate calibration is not critical.

An additional advantage of having the accelerometer attached to the active camera is that we are able to orient the axis of sensitivity to a given position relative to the wearer's body, therefore increasing by robotic means the sensing range. Note that since we use a 2D axis accelerometer, only elevation and cyclotorsion have inertial sensing. The reason to use accelerometers is that they provide an absolute measurement with respect to external frames of reference such as earth's surface. To provide feedback to the pan axis, a 1D gyroscope or visual servoing can be used.

3.5 Controller

For this prototype platform we use a controller embedded in microprocessor's software. This allow us to explore different control strategies in a flexible way, in the knowledge that this imposes bandwidth restrictions that can be relaxed when dedicated circuitry is used. A 16-bit micro-controller is used to interface the actuators and the inertial sensor with the host computer. However the architecture developed allows the motors and accelerometer to be managed by the micro-controller alone without computer intervention. The embedded controller has a digital PID filter which produce

a control signal of the form:

$$u(k) = u(k-1)+K_1e(k)-K_2e(k-1)+K_3e(k-2) \quad (2)$$

with

$$e(k) = |S_s(k) - S_c(k)| \quad (3)$$

denoting the error between the demand or set-point S_s and the current signal S_c at time k. In the case that we are interested in, aligning the robot head with a pre-defined angle against gravity, the set-point should be previously determined (Eq. 1). K_1, K_2 and K_3 are the controller gains.

Figure 3 shows the general control strategy implemented for each of the mechanical degrees of freedom (note that for panning the inertial sensor is not present), with the principal control paths denoted A and B. For path A, the controller programmed in the micro-processor takes feedback from the inertial sensor to control the actuator. This is useful when fulfilling a demand based in world coordinate frames such as in alignment with respect to earth's surface. On path B, the system directly controls the motor's relative position with respect to user's body. The system can therefore switch between the three main frames of reference. For example demands relative to the user follow path B alone (neither inertial nor visual feedback), whereas demands relative to the world follow path A (either inertial or visual feedback) as do demands relative to an independent object (both inertial and visual information).

3.6 Wearability

As mentioned earlier the prototype wearable robot (Figure 1) is worn on the shoulder at intersection of the coronal and left paramedial anatomical planes. The actuators are linked to a collar made out of thermoformed styrene which has a horseshoe form resting on the neck's base. The collar is connected to the host computer and interface microprocessor using a ribbon cable that runs at the back. This kind of cable has the advantage for wearability of a low profile and good bending properties, but signals should be located carefully to avoid cross-talk interference. The total mass of the camera assembly with sensors and actuators is slightly less than 60 gm.

Since the robot's location tends to the medial plane, movements of the shoulder interfere less with the camera motion, and we find in use that the position impairs neither arm motion nor the carrying of rucksacks.

4 Example Results

4.1 Gaze direction and image stabilization

This experiment aims to maintain the camera horizontal and vertical sufficiently so that visual processing can complete the task. Here, elevation and cyclotorsion axes are receiving feedback from the accelerometer. The robotic head is controlled by the microprocessor alone hich receives feedback from the inertial sensors. Results are shown in Figures 4 and 5.

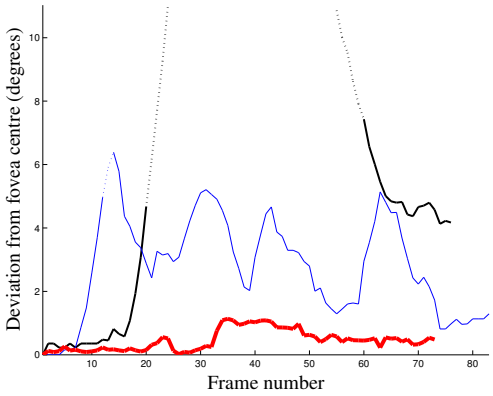

Figure 4: The cooperation between robotic and a software-based virtual fovea improves the image stabilization dramatically. The uppermost curve shows the displacement of the target with no stabilization whatsoever. The middle curve shows that obtained with inertial stabilization alone, and the low curve corresponds to the displacement of target within the virtual moving fovea as shown in Figure 5.

Although the robotic stabilization reduces image motion, it is unable to deal with translation. To minimize in-plane motion we have developed a virtual foveal window as shown in Figure 5. This moving fovea has 3 D.O.F. two for translation and one for cyclotorsion. When the camera translation is small between successive frames, the camera motion is well approximated by a rotation, and a planar homography can be used to transfer a fixation point between images. Planar homographies between pairs of images are calculated here using a point based method and the RANSAC algorithm [18], although other feature and direct methods can be equally applicable for this task.

The result is a moving rectangle inside images that follows the objects centered at the beginning of the sequence (middle row Figure 5). Therefore, the stabilization/tracking of an object is a mixture of the compensation provided by the active sensor, inetrial information and the virtual fovea.

4.2 Visual Servoing

The position of the virtual fovea provides feedback information that helps the active robot to keep objects in the centre of the image. Since the camera calibration is known in advance, it is straightforward to convert image displacements into signals which re-centre the fovea. This task is accomplished with assistance from the inertial sensors since the host computer calculate and sent to the microprocessor the value of S_s in eq. 3. Figure 6 shows some im-

Figure 5: Comparable samples from sequences obtained while the user sits down in an outdor scene (bottom row). The top row shows the view and fovea when the camera is passive, the second row when it is active and the third row for the configuration with active plus software-controlled fovea. Note that the passive camera loses the initial target point for about half the sequence.

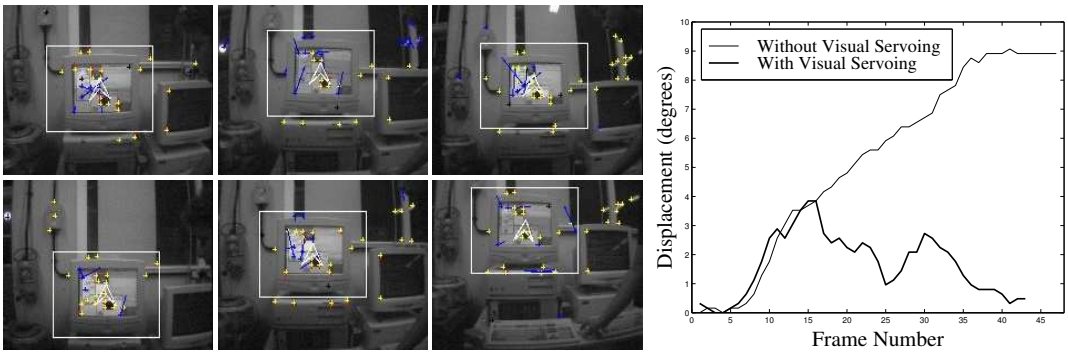

Figure 6: Left: first, middle and last images from a sequence as the user performs an almost pure vertical displacement in front of a close object. In the bottom row, the tilt information alone is unable to maintain fovea centered. The top row uses a fusion between visual and tilt information to maintain the fovea under control. The graph for the full sequences are shown at the right.

agery drawn from sequences with and without visual feedback.

4.3 Saccades

In order to control a saccade (fast redirections of gaze using pre-computed trajectories) we use path B in Figure 3 and neither inertial nor visual information is taken into account. Figure 7 shows images from the WeRo as it performs a panning saccade of about 25° (half of the FOV of the visual sensor). This movement brings a point in the periphery of the FOV to the centre of the image. The redirection is completed in about 2 frames, using a frame-grabber capturing at 25 Hz on a 500 MHz Pentium computer. This performance is comparable to the one achieved by one of our highly engineered active vision heads [19]. Note that vision is of little use during the saccade, as the images suffer motion blur.

4.4 The world as memory

The active vision paradigm encourages the use of compact and 'just in time' representations of the environment. Rather than maintaining dense maps of the surroundings, the gaze of the system can be re-directed to the salient part of the scene.

Instead of having to store prior knowledge of how entire scenes look, the system may try to recognize the context by trying to recognize individual canonical objects and their relative placement (eg [20]).

A small taste of what can be achieved towards world as memory sensing when inertial information is used is shown in Figure 8. The bottom row shows the images from the visual sensor when commanded to switch to a world-based task which in this case is to point to the ceiling. It

Figure 7: Complete sequence during a panning saccade.

Figure 8: Images showing the switching between world and user frames of reference for different postures of the user. Accomplishing this task is enormously simplified by the fusion of inertial information and an active sensing approach.

achieves this regardless of the posture of the user which differs between columns. Information coming from pictures oriented relative to gravity vector may be used as an additional clue for specific room recognition (eg., using color/texture of the floor, ceiling, etc).

5 Conclusions and Discussion

This paper has presented a test prototype for wearable active vision sensor and shown its operation in a coupling-decoupling process of the camera movement from the wearer's posture and movements. It combines an active sensing approach, inertial information and visual sensor feedback. The issues of sensor placement, robot kinematics and their relation with wearability were discussed and the performance of the prototype head has been evaluated in some core visual tasks. Our future work include the evaluation of wearable visual robots with more degrees of freedom, exploration of potential applications in human-human communication as in tele-embodiment and tele-presence, as well as the development of algorithms to deal with the core question of attentional focusing: what should be looked at and how it should be looked from moment to moment?

Acknowledgements

WWM gratefully acknowledges the receipt of Mexican Government CONACYT scholarship. This work and BJT are funded by Grants GR/L58668 and GR/N03266 from the UK's Engineering and Physical Science Research Council.

References

[1] R. Bajcsy, "Active Perception," *Proc.IEEE*, vol. 76, no. 8, pp. 996–1005, 1988.

[2] D. H. Ballard and C. M. Brown, "Principles of animate vision," *CVGIP: Image Understanding*, vol. 56, no. 1, pp. 3–21, 1992.

[3] J. Aloimonos, I. Weiss, and A. Bandyopadhyay, "Active vision," in *1st International Conference on Computer Vision, London*. 1987, pp. 35–54, IEEE Computer Society Press.

[4] Blake and Yuille, *Active Vision*, MIT Press, Cambridge MA, 1992.

[5] T. Starner J. Weaver and A. Pentland, "Real-time american sign language recognition using desk and wearable computer based video," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, December 1998.

[6] B. Schiele and A. Pentland, "Attentional objects for visual context understanding," Tech. Rep. 500, MIT media Lab, 1999.

[7] N. Kohtake J. Rekimoto and Y. Anzai, "Infostick: an interaction device for inter-appliance computing," in *Proc. Workshop on Handheld and Ubiquitous Computing (HUC'99)*, 1999.

[8] H. Aoki B. Schiele and A. Pentland, "Realtime personal positioning system for a wearable computers," in *Proc. International Symp. on Wearable Computing*, 1999.

[9] S. Mann, "Wearcam (the wearable camera)," in *IEEE Int. Symp. on Wearable Computing*, 1998.

[10] J. Healey and R. Picard, "Starlecam: A cybernetic wearable camera," Tech. Rep. 468, MIT Media Lab perceptual Computing section, October 1998.

[11] S.K. Nayar and V Peri, "Folded catadioptric cameras," in *Computer Vision and Pattern Recognition, Fort Collins CO, June 1999*, Los Alamitos, CA, 1999, pp. 217–223, IEEE Computer Society Press.

[12] D. Grant and V. Hayward, "Constrained Force Control of Shape Memory Alloy Actuators," in *IEEE ICRA 2000.*

[13] K. Takemura and T. Maeno, "Characteristics of an Ultrasonic Motor Capable of Generating a Multi-Degrees of Freedom Motion," in *IEEE ICRA 2000.*

[14] A.R. Golding and N. Lesh, "Indoor navigation using a diverse set of cheap wearable sensors," in *Proc. International Symp. on Wearable Computing*, 1999.

[15] J. Farringdon A.J. Moore N. Tilbury J. Church and P.D. Biemond, "Wearable sensor badge and sensor jacket for context awareness," in *Proc. International Symp. on Wearable Computing*, 1999.

[16] H. Bussmann P. Reuvenkamp P. Veltnik et. al, "Validity and reliability of measuerments obtained with an "activity monitor" in people with and without a transtibial amputation," *Physical Therapy*, vol. 78, no. 9, September 1998.

[17] T. Vieville and O. Faugueras, "Computation of inertial information on a robot," in *Fifth Int. Symposium on Robotics Research*, Hirofumi Miura and Suguru Arimoto, Eds. 1989, MIT-Press.

[18] P.H.S. Torr, *Motion segmentation and outlier detection*, Ph.D. thesis, University of Oxford, Dept of Engineering Science, 1995.

[19] D W Murray K J Bradshaw P F McLauchlan I D Reid and P M Sharkey, "Driving saccade to pursuit using image motion," *International Journal of Computer Vision*, pp. 205–228, 1995.

[20] R. D. Rimey and C. M. Brown, "Control of Selective Perception Using Bayes Nets and Decision Theory," *International Journal of Computer Vision*, vol. 12, no. 3, pp. 173–207, Apr. 1994.